Università Iuav di Venezia

Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

2016 Purdue Conferences
Compressor Engineering
Refrigeration and Air Conditioning
High Performance Buildings

FROM ENERGY SIGNATURE TO CLUSTER ANALYSIS: AN INTEGRATED APPROACH

LORENZA PISTORE
GIOVANNI PERNIGOTTO
FRANCESCA CAPPELLETTI
PIERCARLO ROMAGNONI
ANDREA GASPARELLA

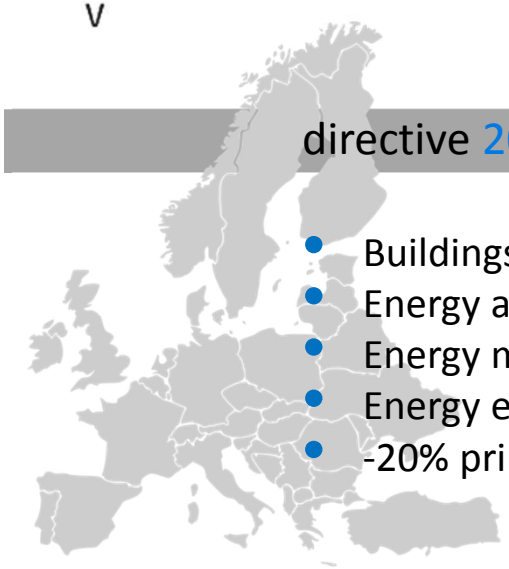
4TH INTERNATIONAL HIGH PERFORMANCE BUILDINGS CONFERENCE
PURDUE UNIVERSITY, JULY 11, 2016

Herrick Laboratories PURDUE UNIVERSITY

PURDUE UNIVERSITY

Center for High Performance Buildings at Purdue

- Introduction
- Objective and case study
- Method
- Results
- Conclusions
- Future outlooks



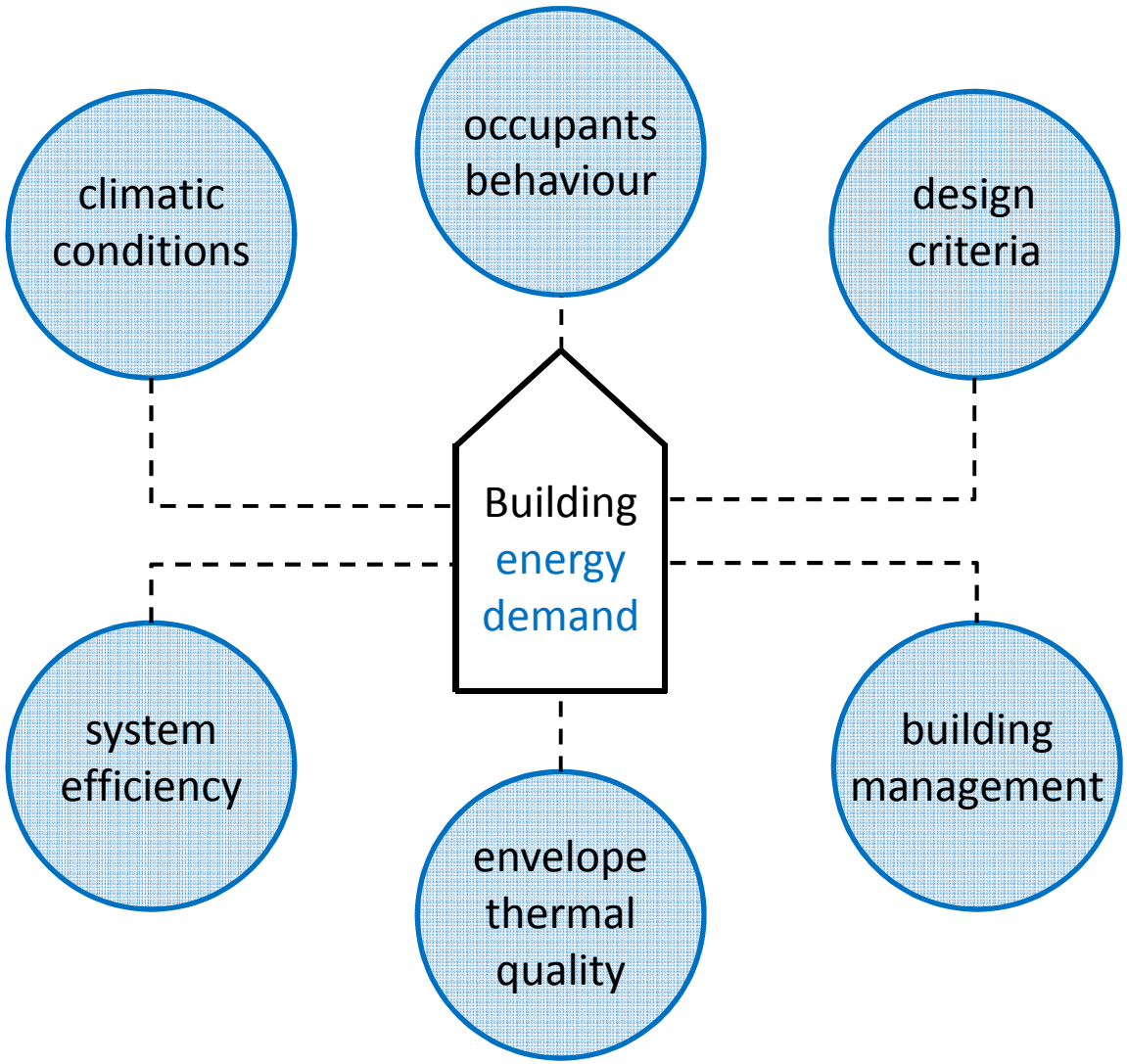
directive 2012/27/EU → Public buildings

- Buildings refurbishment
- Energy audit
- Energy monitoring
- Energy efficiency plans
- -20% primary energy

starting from 2014 there is the obligation to refurbish at least 3% per year of the total surface of public buildings

medium and long-term interventions and operational strategies in order to improve buildings energy efficiency, identifying the cost-optimal ones

in order to identify the actions needed to enhance buildings energy performance and the control strategies, it is important at a first step to define a method of energy analysis and diagnostics



case by case energy diagnosis of a large stock of buildings is onerous :

- extensive monitoring campaigns
- field surveys
- energy performance calculation

How to perform an energy audit of a large stock of buildings ?

objective

starting from a large stock of buildings, grouping them based on their similar features and find a reference building for each cluster

method

1. energy signatures
2. multiple regression model
3. cluster analysis and validation

case study

42 high school buildings

north- east of Italy





1. ENERGY SIGNATURES

ANNEX B of EN 15603:2008
Energy Signatures

A simplified method, suitable for large stock of buildings:

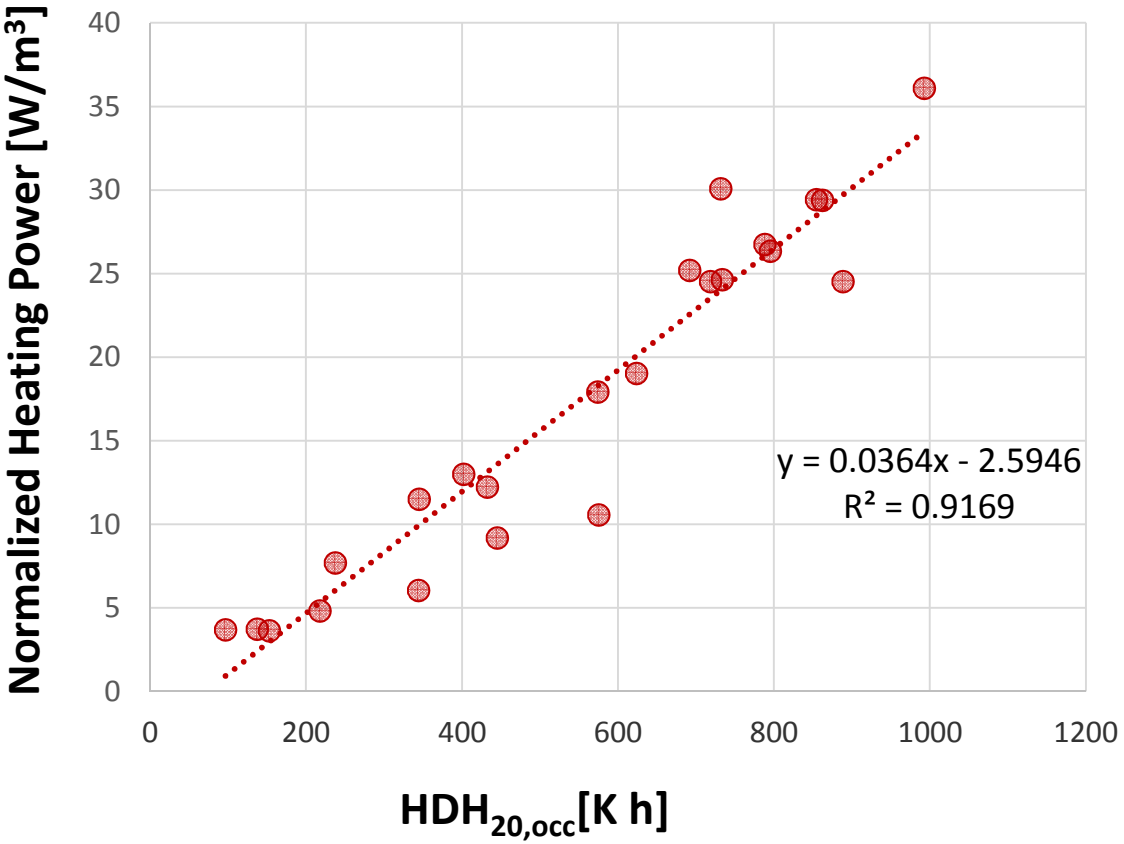
heating and cooling energy demands of a given building are correlated to climatic data over a suitable period

preferably applied for buildings with stable internal gains and relatively low passive solar gains

useful information on building energy performance and allows :

- fast detection of malfunctions or changes in the building operation/features
- the verification of the efficacy of any retrofit interventions

Energy Signature



Normalized Heating Power [W/m3]

the average weekly power per unit volume, obtained by dividing the energy use during one week per unit volume by the amount of opening hours per week

$$\phi = \sum_{i=1}^{7} \frac{EP_{h_i}/V}{\tau}$$

$HDH_{20,occ}$[K h]

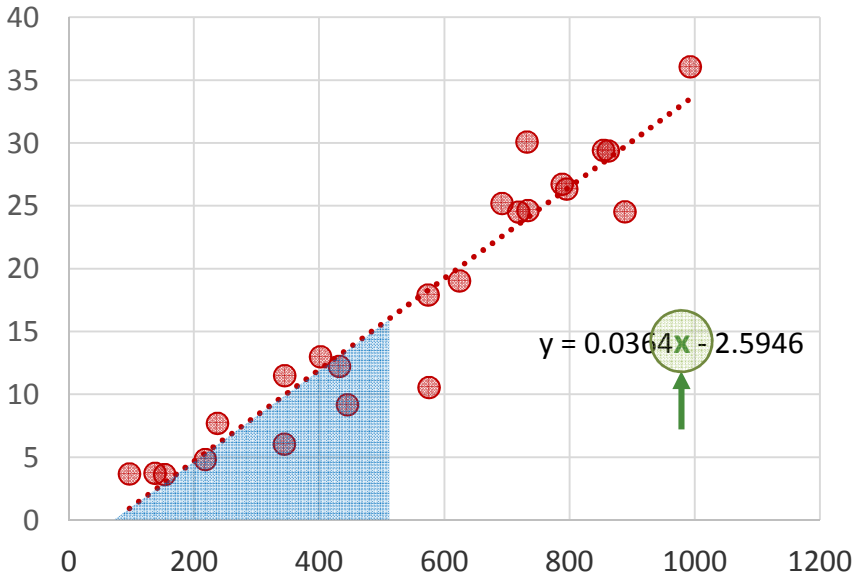
the weekly heating degree-hours during the opening hours

$$HDH_{20,occ} = \sum_{i=1}^{n} (20 - T_{ext})_i$$

n = opening hours of a week

2. MULTIPLE REGRESSION MODEL

Energy signature



zero of the function

represents the minimum number of $HDH_{20,occ}$ for which the system has to be turned on:
the higher is the intercept, the better is the passive building thermal performance

slope

represents the energy performance of the building:
the steeper the slope, the larger is the heating power needed

dependent variables

correlation with some building parameters

12 descriptive parameters

DESIGN CRITERIA

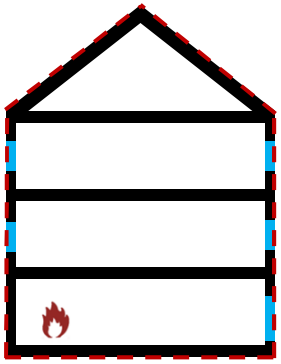
- roof area
- ground floor area
- floor area
- vertical walls area
- windows to walls ratio
- windows to floor ratio
- compactness ratio S/V

ENVELOPE CHARACTERISTICS

- transparent envelope
- opaque envelope area
- transparent to opaque envelope ratio
- average envelope U value

HEATING SYSTEM

- heating system capacity



normalized using the highest value in the whole dataset

multiple linear regression :

find the sets of the candidate quantities which better define homogenous groups

best number of predictors which allow the highest R^2_{adj}

top 10 combinations

clustering

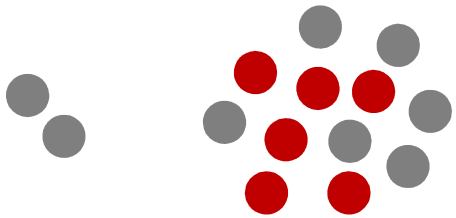
statistical significance?
R^2_{adj}
F-tests
p-values

1 dipendent variables
zero of the funcion
slope

12 predictors

2 predictors

12 predictors

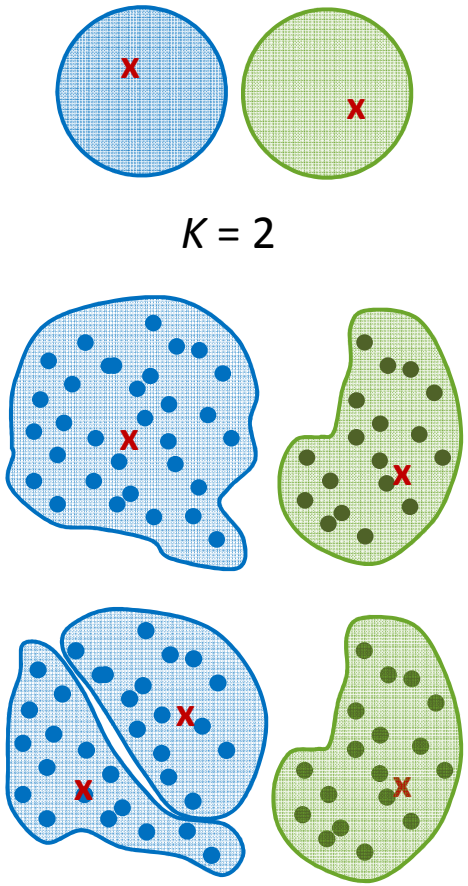


best ID configurations

3. CLUSTER ANALYSIS AND VALIDATION

the *K*-means approach :

a simple partitional algorithm that tries to find *K* non-overlapping clusters



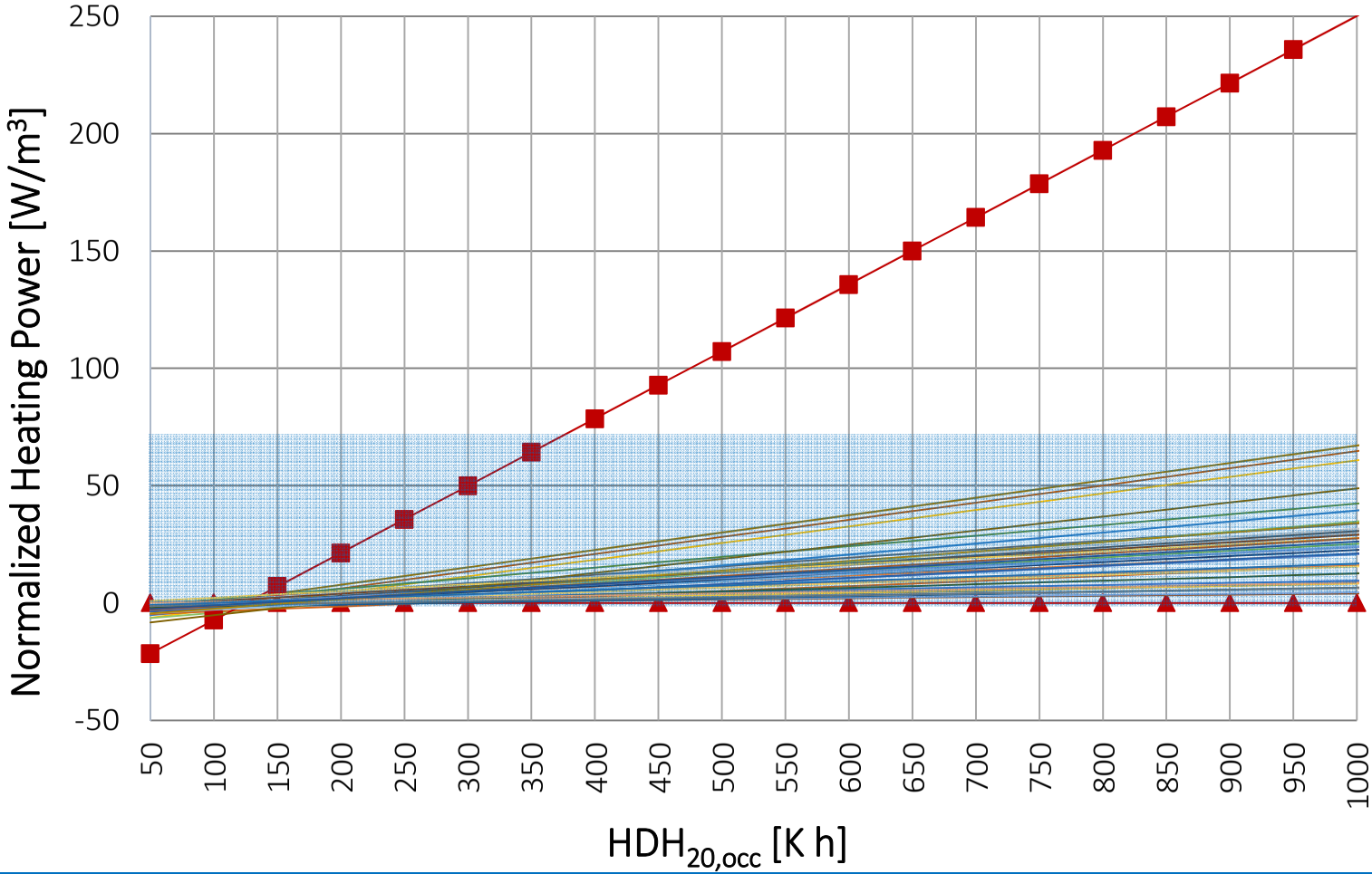
K = 2

- definition of the desired number of clusters
- randomly identification of the virtual centroids
- data points (schools) distribution according to calculation of the distance to the closest centroid
- checking if the combination of predictors with the highest R^2_{adj} previously chosen with respect to the whole dataset is the best for the cluster as well

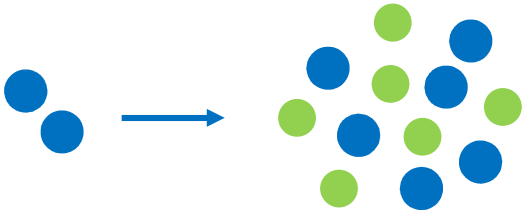
if not ✗

- the combination of predictors with the highest R^2_{adj} is found with another regression and used as new coordinate system, considering to split the clusters

1. ENERGY SIGNATURES



2. MULTIPLE REGRESSION MODEL



6 predictors → highest R^2_{adj}

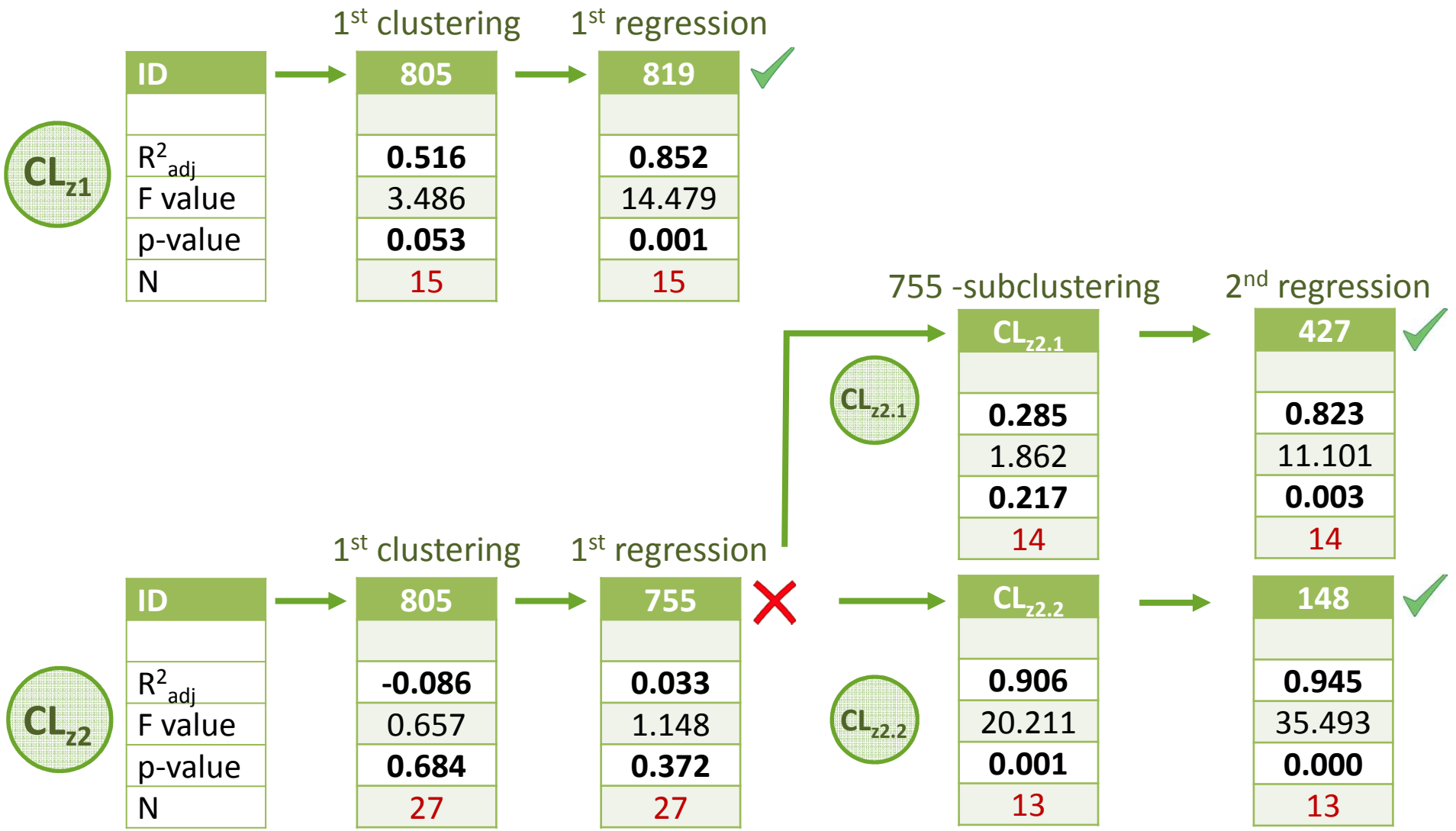
924 combinations → top 10 configurations

Zero of the function										
ID	**805**	**286**	**861**	**346**	**356**	**614**	**871**	**851**	**872**	**362**
R^2_{adj}	0.16	0.15	0.15	0.15	0.15	0.14	0.14	0.14	0.14	0.14
F value	2.27	2.24	2.23	2.17	2.16	2.16	2.15	2.15	2.11	2.11
p-value	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.08	0.08

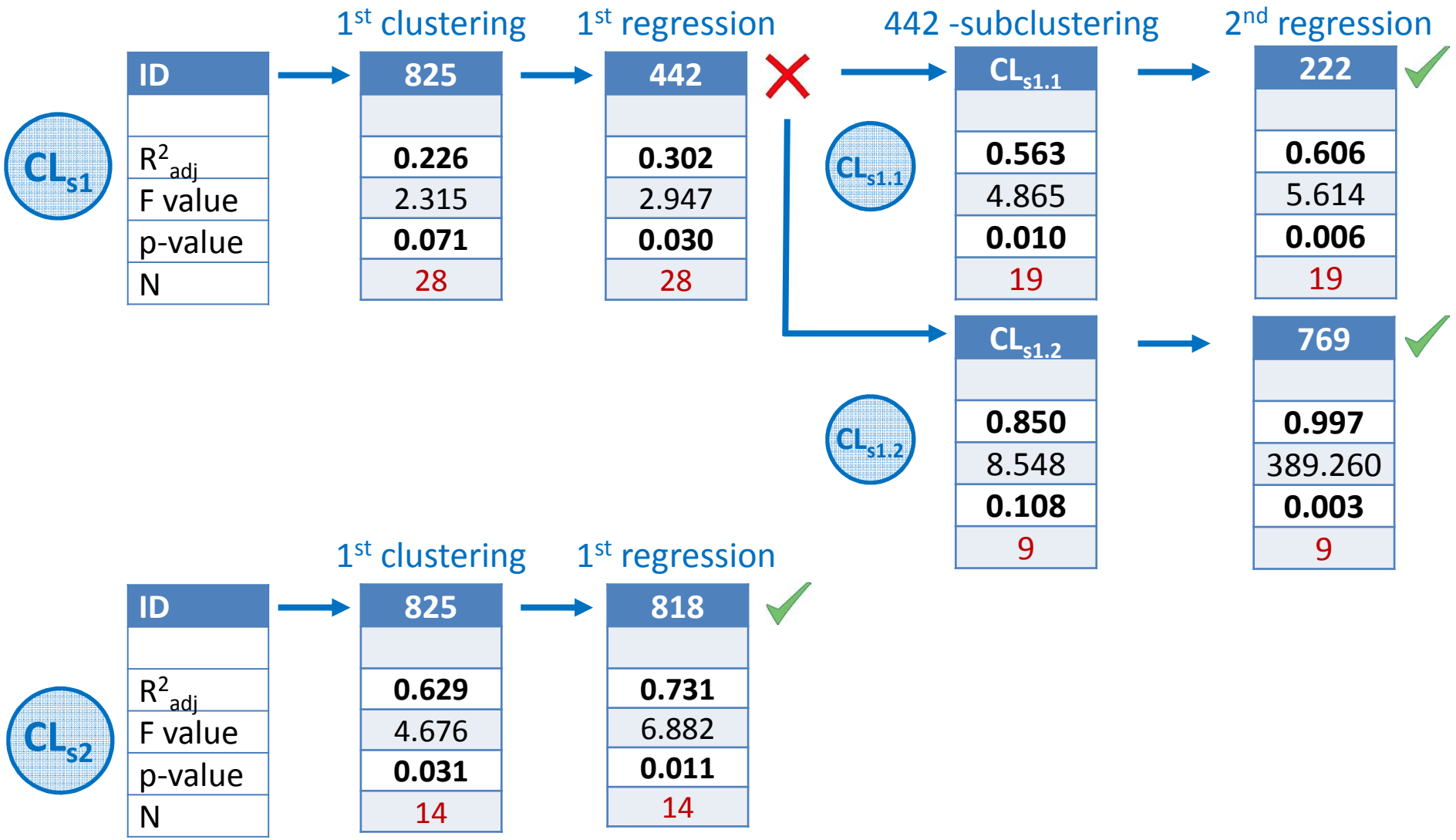
Slope										
ID	**902**	**557**	**662**	**761**	**661**	**822**	**627**	**881**	**825**	**762**
R^2_{adj}	0.17	0.16	0.16	0.15	0.15	0.15	0.15	0.15	0.15	0.15
F value	2.38	2.34	2.28	2.23	2.22	2.21	2.20	2.20	2.19	2.18
p-value	0.05	0.05	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07

3. CLUSTER ANALYSIS AND VALIDATION

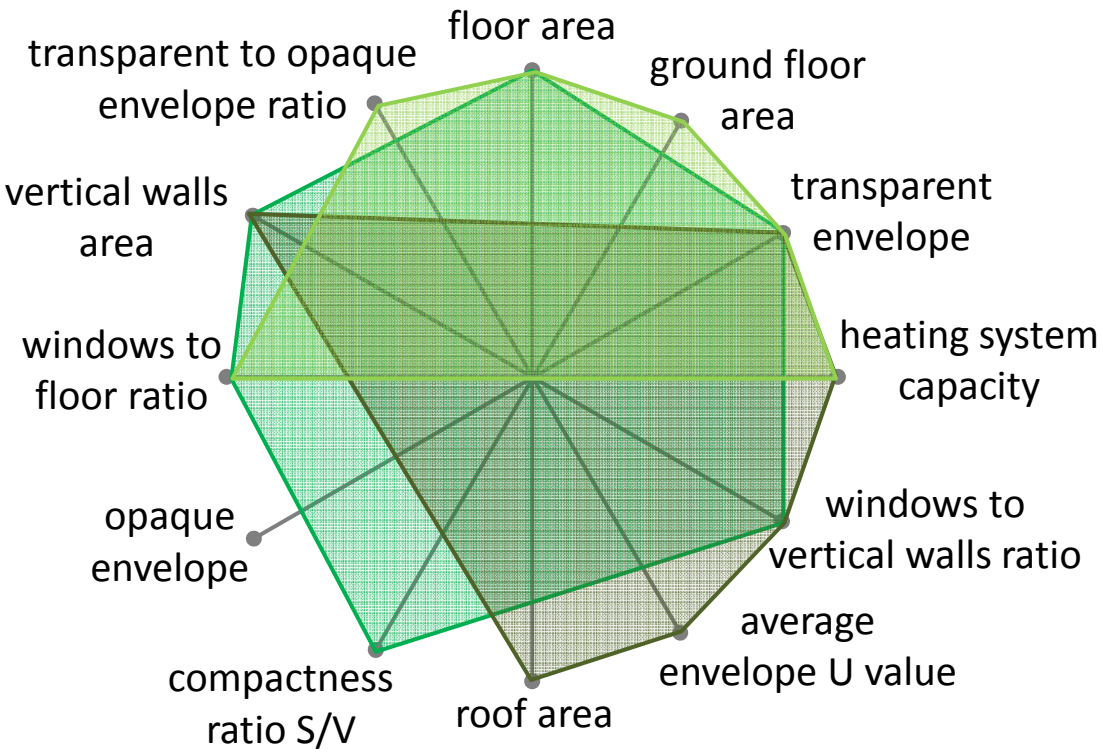
		Zero of the function									
ID		**805**	**286**	**861**	**346**	**356**	**614**	**871**	**851**	**872**	**362**
CL_{z1}	R^2_{adj}	**0.516**	0.322	0.121	0.558	0.100	0.225	-0.282	-0.011	-0.006	0.004
	F value	3.486	1.870	1.734	3.107	1.558	1.339	0.707	0.941	0.966	1.020
	p-value	**0.053**	0.254	0.153	0.146	0.203	0.579	0.686	0.483	0.466	0.434
	N	**15**	12	33	11	31	8	9	34	34	34
CL_{z2}	R^2_{adj}	**-0.086**	0.102	0.325	0.066	0.286	0.006	0.012	0.597	0.836	0.255
	F value	0.657	1.550	1.641	1.355	1.666	1.034	1.067	2.730	6.937	1.399
	p-value	**0.684**	0.207	0.426	0.273	0.323	0.425	0.407	0.433	0.283	0.570
	N	**27**	30	9	31	11	34	33	8	8	8


1st clustering
1st regression
ID
805
819
CLz1
R2adj
F value
p-value
N
0.516
3.486
0.053
15
0.852
14.479
0.001
15
755 -subclustering
2nd regression
CLz2.1
427
0.285
1.862
0.217
14
0.823
11.101
0.003
14
1st clustering
1st regression
ID
805
755
CLz2
R2adj
F value
p-value
N
-0.086
0.657
0.684
27
0.033
1.148
0.372
27
CLz2.2
148
0.906
20.211
0.001
13
0.945
35.493
0.000
13

		Slope									
ID		**902**	**557**	**662**	**761**	**661**	**822**	**627**	**881**	**825**	**762**
CL_{s1}	R^2_{adj}	0.109	0.200	0.091	0.249	1.00	0.020	0.117	0.141	**0.226**	0.522
	F value	1.388	1.960	1.469	2.772	-	1.078	1.732	1.929	2.315	2.640
	p-value	0.290	0.129	0.234	0.032	-	0.413	0.152	0.111	**0.071**	0.228
	N	20	24	29	33	1	24	34	35	**28**	10
CL_{s2}	R^2_{adj}	0.489	0.566	-0.060	0.896	0.148	0.014	0.763	-	**0.629**	0.149
	F value	4.349	4.701	0.886	12.484	2.155	1.040	4.753	-	4.676	1.906
	p-value	0.010	0.013	0.556	0.076	0.072	0.450	0.337	-	**0.031**	0.119
	N	22	18	13	9	41	18	8	7	**14**	32


1st clustering
1st regression
442 -subclustering
2nd regression
CLs1
ID
R2adj
F value
p-value
N
825
0.226
2.315
0.071
28
442
0.302
2.947
0.030
28
CLs1.1
0.563
4.865
0.010
19
222
0.606
5.614
0.006
19
CLs1.2
0.850
8.548
0.108
9
769
0.997
389.260
0.003
9
1st clustering
1st regression
CLs2
ID
R2adj
F value
p-value
N
825
0.629
4.676
0.031
14
818
0.731
6.882
0.011
14

involved predictors: zero of the function



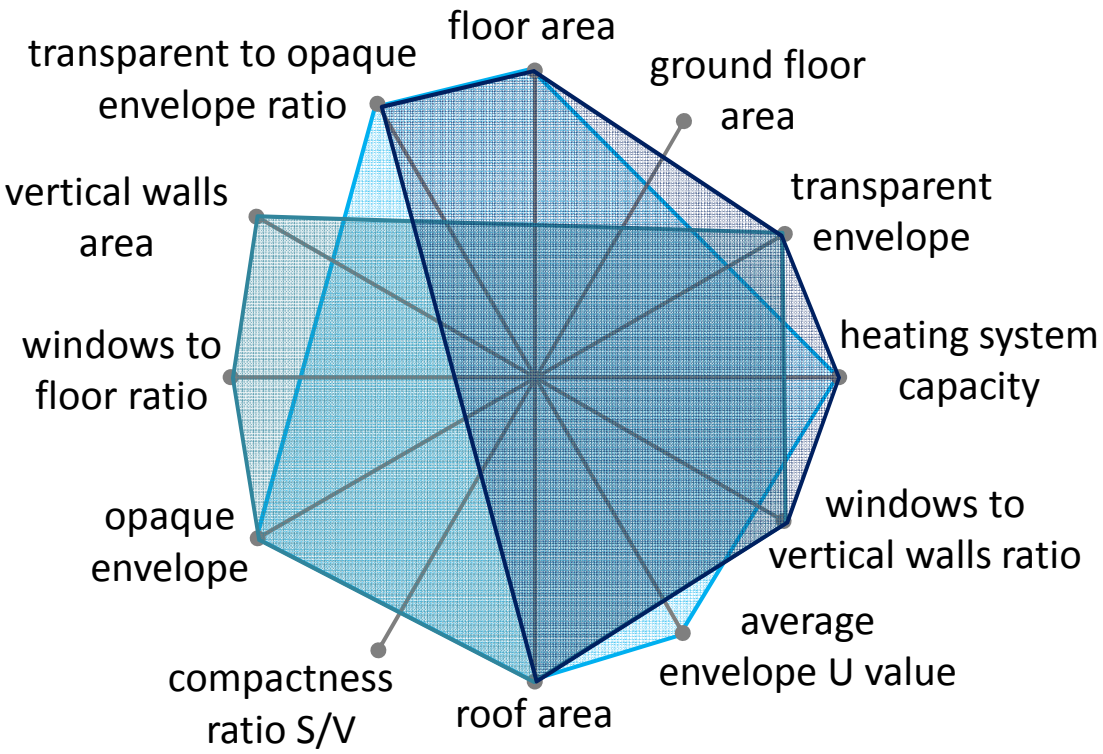
schools identified as cluster centroids

CL_{z1} : CN042_01

$CL_{z2.1}$: TV032_01

$CL_{z2.2}$: VV127_01

involved predictors: slope



schools identified as cluster centroids

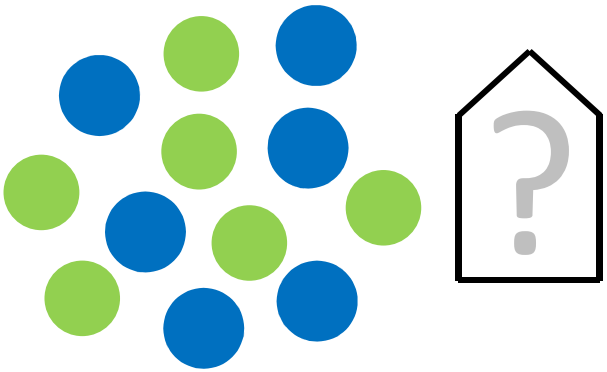
CL_{s2} : TV116_01

$CL_{s1.1}$: MB083_02

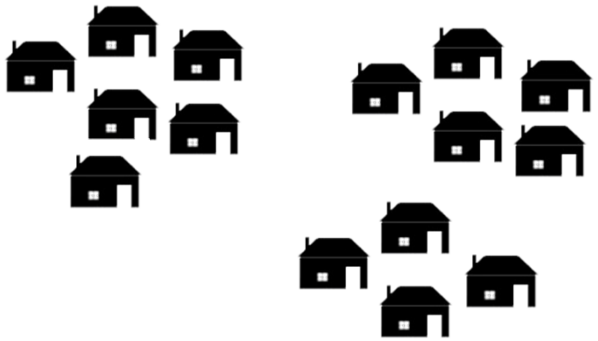
$CL_{s1.2}$: CN042_01

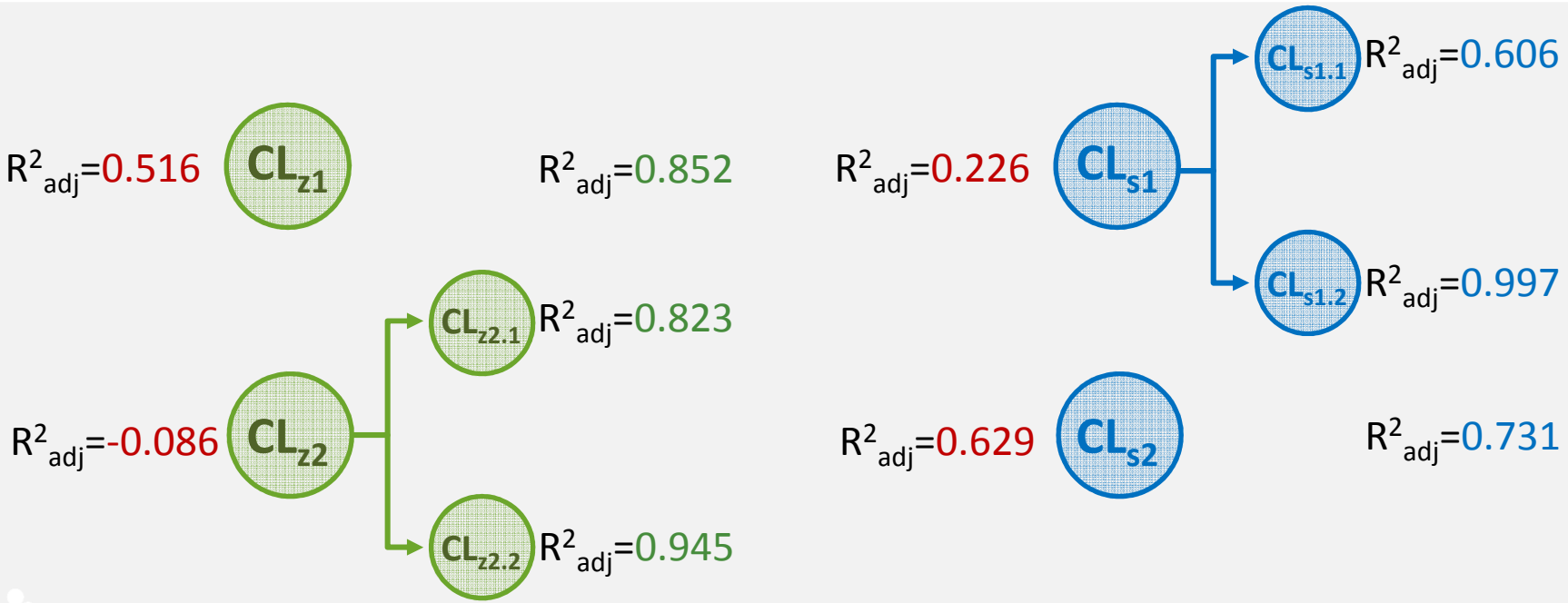
- Starting from few data about a large building stock, using the energy signature method it has been possible to analyze and compare buildings energy performance without onerous and long-term monitoring campaigns

- Starting from 12 descriptive variables, correlated with the normalized zero of the functions and the normalized slopes of the outcome energy signatures, it has been possible to highlight 6 most influencing predictors



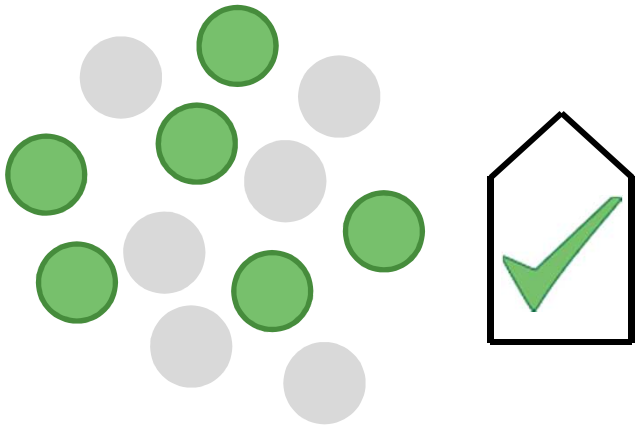
- After the identification of the best group of predictors, a cluster analysis has been performed and validated
- Data in the clusters have been studied and optimized with regressions, obtaining a good improvement in the R^2 values





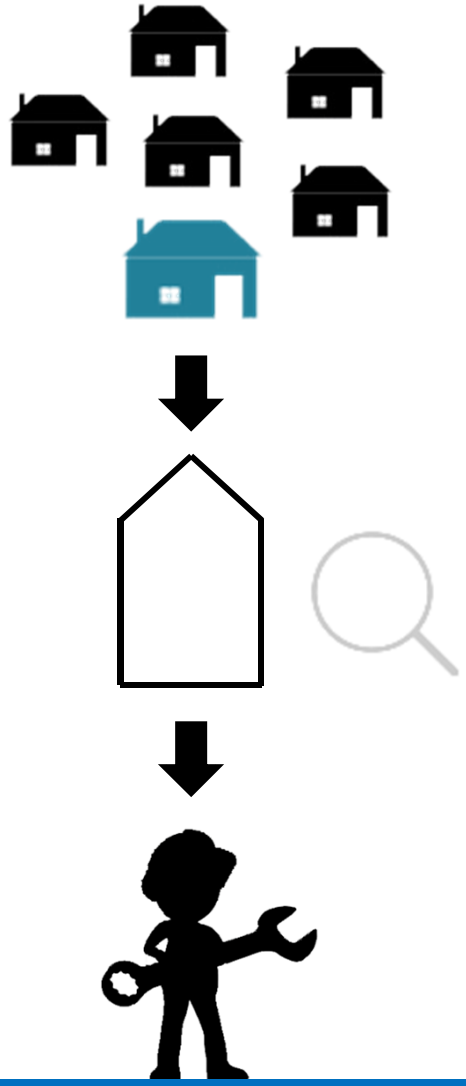
- For each cluster, the reference building has been identified

- For each final best ID configuration, it has been possible to select the 6 predictors that describe the set of school included in the clusters, which can be, in additional phase, the areas of intervention to act on



- Taking into account the set of variables most influencing each cluster of schools, it will be possible, using a cost-optimal approach, to list the potential interventions, starting from the reference building and extend these solutions to the other schools in the same cluster

- Conducting the *ES approach* also in the further years, there is the possibility to check the potentialities of this method to highlight the effects of performed energy efficiency measures



THANKS FOR YOUR ATTENTION!

lorenza.pistore@natec.unibz.it